

央行金融文本情绪预测方法步骤

1. 打开央行货币政策报告官网（下载季度货币政策报告）

<http://www.pbc.gov.cn/zhengcehuobisi/125207/125227/125957/126027/2895896/index.html>

2. 将 PDF 报告转成 WORD 格式，方便后续分词处理

3. 进行文本情绪分析用到的软件：Matlab python

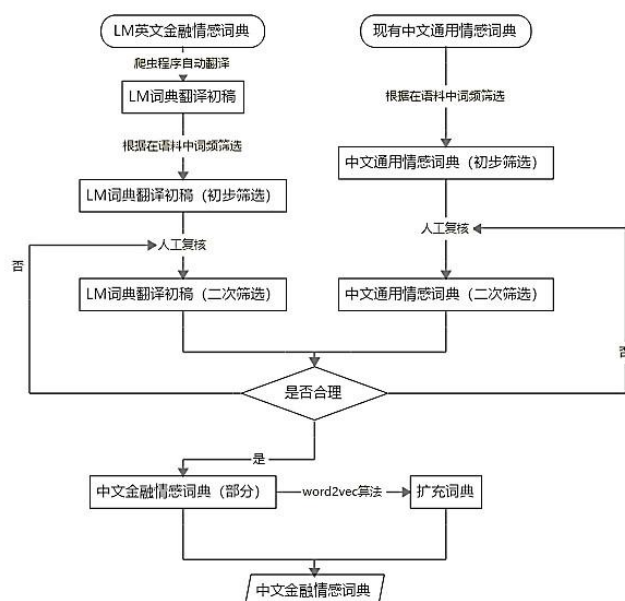
4. 在进行文本情绪分析时需有已进行词语分类的中文金融词典库进行对照，中文金融情感词典见 excel(参考姜富伟构建的词典)

中文金融情感词典参考链接

https://www.cnopendata.com/data/chinese-dictionary-of-financial-sentiment?module=public_data

注：中文金融情感词典构建原理及方法

在 Loughran and MacDonald（2011）词典（英文金融情感词典）的基础上通过人工筛选和 word2vec 算法扩充,构建了一个更新更全面的中文金融情感词典。我们使用该情感词典计算我国财经媒体文本情绪指标,发现媒体文本情绪可以更准确地衡量我国股市投资者情绪的变化,对我国股票回报有显著的样本内和样本外预测能力。



5. 高级金融文本情绪分析方法-基于语句的分析

情绪分析流程参考链接: <https://jishuin.proginn.com/p/763bfd60498>

中文情感词典下载链接:

<https://pan.baidu.com/s/1RjFJzehEHGyCLbC7GtCYWQ> 密码:esdp

(1) 词典标注思路

- **正式词典标注思路:** 正式用语情感词典, 通过年报公布后 3 个交易日累积正负收益率为标准, 将年报标记为正负面情绪两类。
- **非正式词典标注思路:** 使用所有中国上市公司在雪球论坛和东方财富股吧内相关帖子, 共 8130 万条。在网络股票论坛, 用户发表自己的意见时, 经常带有表情符号, 从而使得帖子带有明显的情绪指标。这种含有特殊指标的帖子, 省去了人工标注文本情绪的工作。

论文已经公开了中文情感词典, 将其整理为 4 个 txt 文件

formal_pos.txt 正式用语正面情绪词典

formal_neg.txt 正式用语负面情绪词典

unformal_pos.txt 非正式用语正面情绪词典

unformal_neg.txt 非正式用语负面情绪词典

(2) 中文金融词典使用方法 (python)

- 年报正式用语词典

```
from cnsenti import Sentiment

senti = Sentiment(pos='dict/formal_pos.txt', #正面词典txt文件相对路径
                  neg='dict/formal_neg.txt', #负面词典txt文件相对路径
                  merge=False,               #是否将cnsenti自带词典和用户导入的自定义词典融合
                  encoding='utf-8')         #两txt均为utf-8编码

test_text = '这家公司是行业的引领者, 是中流砥柱。今年的业绩非常好。'
result = senti.sentiment_count(test_text)

print('sentiment_count', result)
```

from cnsenti import Sentiment

senti = Sentiment(pos='dict/formal_pos.txt', #正面词典 txt 文件相对路径

neg='dict/formal_neg.txt', #负面词典 txt 文件相对路径

```

merge=False, #是否将 cnsenti 自带词典和用户导入的自定义词典融合
encoding='utf-8') #两 txt 均为 utf-8 编码
test_text = '这家公司是行业的引领者，是中流砥柱。今年的业绩非常好。'
result = senti.sentiment_count(test_text)
print('sentiment_count',result)
sentiment_count {'words': 16, 'sentences': 2, 'pos': 3, 'neg': 0} #结果显示

```

● 财经社交媒体非正式用语词典

```

from cnsenti import Sentiment

senti = Sentiment(pos='dict/unformal_pos.txt', #正面词典txt文件相对路径
                  neg='dict/unformal_neg.txt', #负面词典txt文件相对路径
                  merge=False,                #融合cnsenti自带词典和用户导入的自定义词典
                  encoding='utf-8')           #两txt均为utf-8编码

test_text = '这个股票前期走势承压，现在阴跌，散户只能割肉离场，这股票真垃圾'
result = senti.sentiment_count(test_text)

print('sentiment_count',result)

```

```

from cnsenti import Sentiment

senti = Sentiment(pos='dict/unformal_pos.txt', #正面词典 txt 文件相对路径
                  neg='dict/unformal_neg.txt', #负面词典 txt 文件相对路径
                  merge=False,                #融合 cnsenti 自带词典和用户导入的自定义词典
                  encoding='utf-8')           #两 txt 均为 utf-8 编码

test_text = '这个股票前期走势承压,现在阴跌,散户只能割肉离场,这股票真垃圾'
result = senti.sentiment_count(test_text)
print('sentiment_count',result)
sentiment_count {'words': 18, 'sentences': 1, 'pos': 0, 'neg': 2} #运行结果

```